LA-UR-21-22988

Approved for public release; distribution is unlimited.

Title:	Evaluation of the Los Alamos Nuclear Material Packaging Risk Ranking Method
Author(s):	Smith, Paul Herrick Kelly, Elizabeth J. Kaufeld, Kimberly Ann Stone, Timothy Amos Prochnow, David Adrian Grow, David Isaac Karns, Tristan Davis, John Taylor Balkey, Simon Narlesky, Joshua Edward
Intended for:	Report
Issued:	2021-04-01 (rev.1)

Evaluation of the Los Alamos Nuclear Material Packaging Risk Ranking Method

Paul H. Smith, Elizabeth J. Kelly, Kimberly A. Kaufeld, Tim A. Stone, David A. Prochnow, David I. Grow, Tristan M. Karns, John T. Davis, Simon Balkey, Joshua Narlesky

1 Introduction

1.1 Why do we have a Risk Ranking Methodology?

Repackaging nuclear material into robust containers to protect workers and the public has been ongoing at LANL and around the DOE complex for nearly two decades. The number of containers at LANL is around 5,000; limited resources for repackaging material has led to extended repackaging campaigns and the need to prioritize repackaging.

Various methodologies have been used to prioritize the repackaging efforts and to demonstrate progress in risk reduction over time (e.g., Boerigter, 1997). The 2000-1 DNFSB recommendation recognized the limited DOE resources for repackaging, and acknowledged the need to "prioritize and schedule tasks to be undertaken with available funds according to consideration of risks." Later, in DNFSB recommendation 2005-1, in addition to recommending that DOE develop a packaging standard, the Board recommended that "Characterization information should also be used to develop a surveillance program prioritized according to expected material and container risk (including, for example, material type, material form, and the age and type of container)."

In response to requests and recommendations from the DOE and DNSFB to prioritize according to worker risk, a risk ranking method based on the potential consequence of dropping a container from 3 meters was developed in 2007 (Smith, 2007) and updated in 2014 (Hoffman, 2014). Various LANL implementation plans for repackaging were developed over the years using this methodology (Stone, 2014). Currently, this method is utilized in conjunction with an algorithm to mitigate programmatic risk to prioritize container repackaging and material processing (Prochnow, 2015). The purpose of this study is to document how the current risk ranking method works, how it is used and potential limitations.

1.2 Why does the risk ranking method need to be evaluated?

Processes sometimes take on a certain momentum and may continue indefinitely if they are not reevaluated with a reasonable frequency. Recognizing the need for continuous improvement, a programmatic milestone was established for March, 2021 to "Complete evaluation and review of legacy container risk ranking methodology." A team of subject matter experts was assembled, with expertise in mathematics, engineering, nuclear material process planning, storage optimization, packaging engineering, and hands-on nuclear materials operations. The team included some of the original authors of the current risk ranking methodology, and it was balanced with new team members who could bring a fresh perspective to the evaluation. This paper documents the results of this evaluation.

1.3 What are the key questions to be evaluated?

Some of the key questions that are addressed in this evaluation include:

- Is a risk-related ranking for container prioritization and reporting still required and/or needed?
- If so, is the current methodology still appropriate - specifically, are the factors, their levels, and the particulars of their combination consistent with additional insights that have been accumulated over the intervening time.
- How is the methodology currently used and is there any reason to reconsider its implementation?
- What are some of the positive and negative attributes of the current method?

2 Description of Risk Ranking Method and Utilization

2.1 Details of the current risk ranking methodology

The current risk ranking methodology is first described in Smith, et al., 2007 (Smith, 2007). A second paper, Hoffman, et al., 2014 (Hoffman, 2014) provides updates to the 2007 paper. The methodology uses a risk ranking metric (*RiskRanking*) that is based on the dose (*Dose*) to a worker from indoor airborne dispersion resulting from a drop of the container from a height of 3 meters times a failure index (*FailureIndex*) that is based on the reactivity of the contents (*ReactivityIndex*), an age estimate for a container (*Age*), and the robustness of the container (*PackageFactor*). *ReactivityIndex* and *PackageFactor* are based on expert judgment. (Note that the 2007 paper did not include *PackageFactor*, it was introduced in 2014.)

RiskRanking is calculated as:

$$RiskRanking = RelativeDose * FailureIndex$$

Dose is defined as:

$$Dose = [DCF]*[ElementWt]*[RRF]*0.0007734, \text{ where}$$

DCF = the 50 year inhalation dose conversion factor = *rem CDE/g* (Table 2 in Smith, et al., 2007)

RRF = the product of the airborne release fraction, the respirable fraction and the damage ratio. (*RRF*s are given for each Item Description Code (IDC) in Table 1 in Smith, et al., 2007 and updated in Hoffman, et al., 2014).

$ElementWt$ = MAR (grams)
0.0007734 is the dose dilution factor

RelativeDose is defined as:

$$RelativeDose = 10 * Dose^{1/6}$$

The *RiskRanking* calculation uses *RelativeDose* rather than *Dose* because *Dose* ranges from less than one to over 10^6 and dominates the other factors in the risk ranking. Logarithms were not taken because zeroes would be introduced. In addition, to quote Kirk Veirs, "this transformation was chosen because lower dose items in poor packaging are risky. Capturing this without losing the importance of high dose was what I wanted to do. And 10^6 versus $5x10^5$ are pretty much equivalent to me. The dose numbers represent potential doses not actual doses. There are so many factors that go into an accident resulting in an actual dose that treating the potential doses as real is not realistic."

The *FailureIndex* of a package is given by

$FailureIndex = Reactivity^2/ReactivityMax * Age/4 * PackageFactor$, where

Reactivity is the sum of the means of values assigned by multiple Subject Matter Experts (SME) to four hazard characteristics: pyrophoricity, corrosivity, pressure and oxidation expansion for each IDC. Note that in the code ReactivityIndex = HotButton(HB).

ReactivityMax = Maximum ReactivityIndex across all IDCs. See Table 3 in Smith, et al., 2007.

Note that in this report

$ReactivityIndex = Reactivity^2/ReactivityMax$, so

$$FailureIndex = ReactivityIndex * Age/4 * PackageFactor$$

Age is the time in years that a container has been in storage. It is determined from data collected from the NMCA database. For most containers the age is determined by subtracting the date a container was first loaded with material from the current date. For older containers, where the date of first material loading is unavailable, the date the material was created is used as a conservative estimate of the material loading date.

PackageFactor is a value ranging from 0 to 1 assigned by SMEs that corresponds to how "robust" the container is, or qualitatively, how close is the container to meeting currently accepted nuclear material packaging design criteria. Containers known as pressure cookers or Mound sample containers are assigned a value of 0.2. The containers known as SNMC's (Standard Nuclear Material Containers, aka Hagan containers), designed to meet an earlier DOE packaging standard (Curtis, 1995) were assigned a value of 0.4. The class of filtered containers called SNMC's includes both threaded top closure stainless steel containers ranging in size from 1-12 qt, and closure ring drums that are either 5 gallons or 10 gallons. The threaded top SNMC's with known nonconforming conditions were assigned values of 0.6 and 0.8 for filter gasket and Tamper Indicating Device (TID) bar weld Nonconformance Reports (NCR), respectively. Package factors are documented in Table 1 of Hoffman, et al., 2014. A PackageFactor = 1 is identified as a non-standard in Table 1 of Hoffman, et al., 2014.

The terms *standard* and *non-standard* when applied to containers essentially align with whether the design satisfies the requirements of DOE M441.1-1 standard (e.g., SAVY's, FSO's, etc.) or equivalent (e.g., 3013's, special form, inside a glovebox, etc.). Understandably, there is, therefore, a strong correlation between *PackageFactor* and these categories. For our purposes, among the containers listed in Table 1 of Hoffman, et al., 2014 those with $0 \le PF \le 0.4$ are *standard*, with one exception, and those with $0.4 < PF \le 1$ are *non-standard*. The one exception is pressure cookers, which are particularly robust (hence PF=0.2), but do not otherwise meet the criteria for *standard.* Additional factors include as a purpose-driven design, the existence of packaging procedures, and sufficient pedigree.

The *RiskRanking* values are grouped into categories using the following scheme (Stone, et al., 2014) :

> 0 and <= 65.5	Low risk
> 65.5 and <= 377	Medium risk
> 377 and <= 666	High risk
> 666	Very High Risk

These categories are based on *RiskRanking* values for the container population in 2014. There is no documentation for how they were determined. Kirk Veirs believes that the boundary between High Risk and Very High Risk was set to capture the Pu 238 items of concern at the time.

The definition of risk is "consequence times the probability of the consequence." In the current risk ranking methodology the consequence is relative dose, but the calculation of the probability of a container failure resulting in the respective relative dose is not based on failure data. Rather, a failure index based on expert judgement is used that includes variables that are likely to impact that failure probability. However, if and how these variables actually impact risk is not fully understood (e.g., are they equally important, or are some more important than others, are there other factors that are equally or more important?). It is unfortunate that in the papers describing the methodology (Smith, et al., 2007, Hoffman, et al., 2014) and that refer to it (Stone, et al., 2014), the methodology is described as using "… a calculated probability of container failure." It is important to recognize that the probability of container failure is not known and the *FailureIndex* is not a probability. *RiskRanking* values are not true measures of risk but are relative ranking values for disposition prioritization. The risk categories are really relative ranking categories driven by subject matter experts, and the thresholds between the categories are somewhat arbitrary. Likewise, the change in the sum of *RiskRanking* values over time for a subset of containers is not a true measure of change in risk for those containers.

2.2 Evaluating Impact of Factors in Risk Ranking

RiskRanking is a function of the variables *Age*, *RelativeDose*, *ReactivityIndex* and *PackageFactor*. Scatterplots of *RiskRanking* versus *Age, RelativeDose* and *ReactivityIndex* grouped by *PackageFactor* are analyzed to better understand how these variables impact *RiskRanking*.

The scatterplot of *RiskRanking* and *Age* by *PackageFactor* is shown in Figure 1. The *PackageFactors* are: 1 (yellow box) non-standard containers, 0.8 (orange cross) NCR weld issue Hagans, 0.6 (orange square) NCR gasket issue Hagans, 0.4 (teal triangles) Hagans, and 0.2 (black dots) pressure cookers or mound sample containers. The Hagans (0.4) are the only group showing an increasing trend with *Age*. The maximum *Age* for the Hagans is approximately 20 years. The non-standards begin at around 20 years of *Age* and show no trend and with a *RiskRanking* varying from almost zero to over 350.

Since *RiskRanking* is a function of *Age*, one wonders about the absence of a trend. Grouping on standard and non-standard containers reduces the impact of the *PackageFactor* and one can evaluate the impact of the other term in the *RiskRanking* equation - *RelativeDose*ReactivityIndex.* The plot of *RelativeDose*ReactivityIndex* versus *Age* for the non-standard containers explains the lack of a trend (Figure 2). This lack of a trend is a result of decreasing *RelativeDose*ReactivityIndex* for the non-standards as a function of *Age*. The same decreasing trend with *Age* is seen in a plot of *RelativeDose* versus *Age*, but it is more dramatic when the *ReactivityIndex* is included. In contrast, there is no trend in *RelativeDose*ReactivityIndex* as a function of *Age* for the standards *(Figure 3).* Therefore the increasing trend in *RiskRanking* versus Age is not eliminated by a decreasing trend in *RelativeDose*ReactivityIndex* for these containers.

The decreasing trend for the non-standards as a function of *Age* suggests some degree of programmatic consistency in reducing the numbers of older containers with high *RelativeDose*ReactivityIndex*.

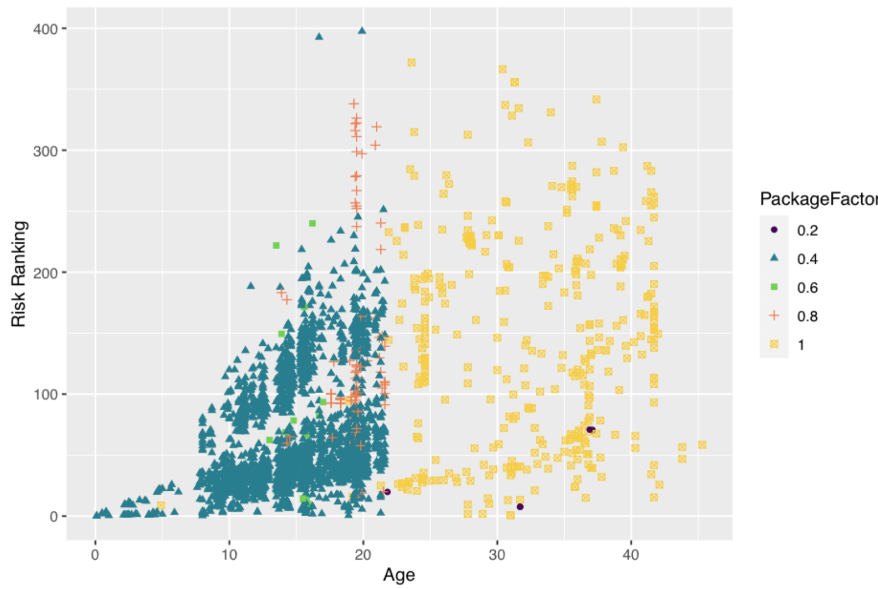

Figure 1. Scatter Plot of *RiskRanking* and *Age* grouped by *PackageFactor*.

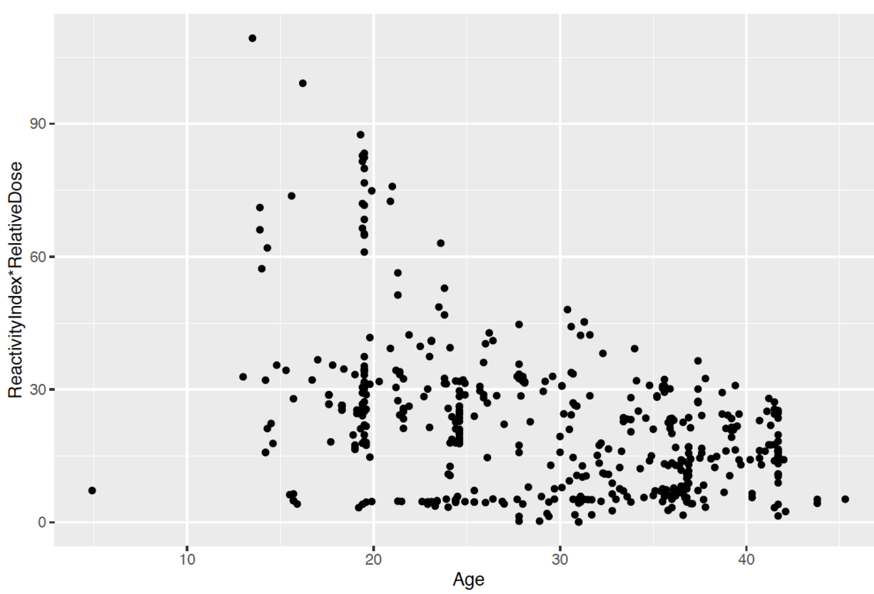

Figure 2. Decreasing *RelativeDose**ReactivityIndex* as a function of *Age* for the non-standards.

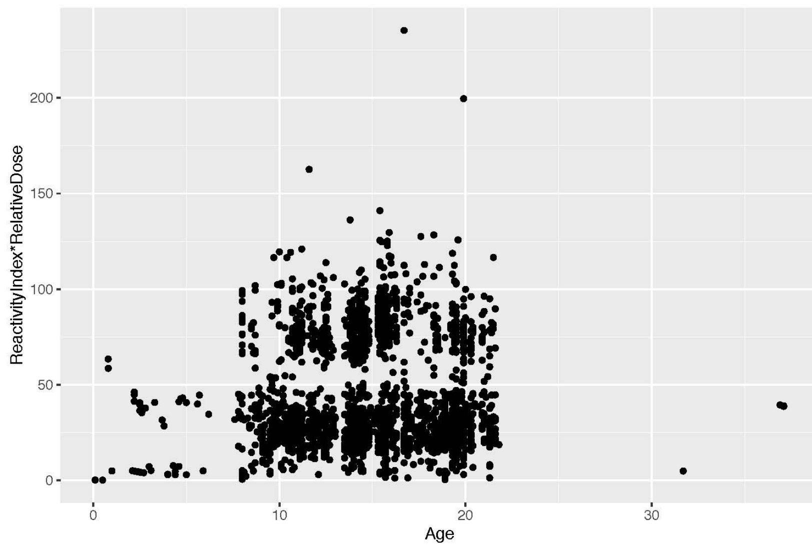

Figure 3. No trend in *RelativeDose*ReactivityIndex* as a function of *Age* for the standards (mainly Hagans).

The scatterplot of *RiskRanking* and *RelativeDose* by *PackageFactor* is shown in Figure 4. There is a clear trend in *RiskRanking* with *RelativeDose* for *PackageFactor* = 1.0. The group with *PackageFactor* = 0.8 (NCR weld issue Hagans) appears to have two subgroups, one with a much steeper trend than the other. These subgroups are due to different IDC's with different *ReactivityFactors*. The group with a higher *RiskRanking* has higher *ReactivityIndex* values (median approximately 3) than the group with lower *RiskRanking* (median approximately 0.9). The *PackageFactor* = 0.4 group (Hagans) also has two subgroups, one with a steeper trend than the other. Again, the group with higher *RiskRanking* has higher *ReactivityIndex* values (median approximately 3) than the group with lower *RiskRanking* (median approximately 0.9).

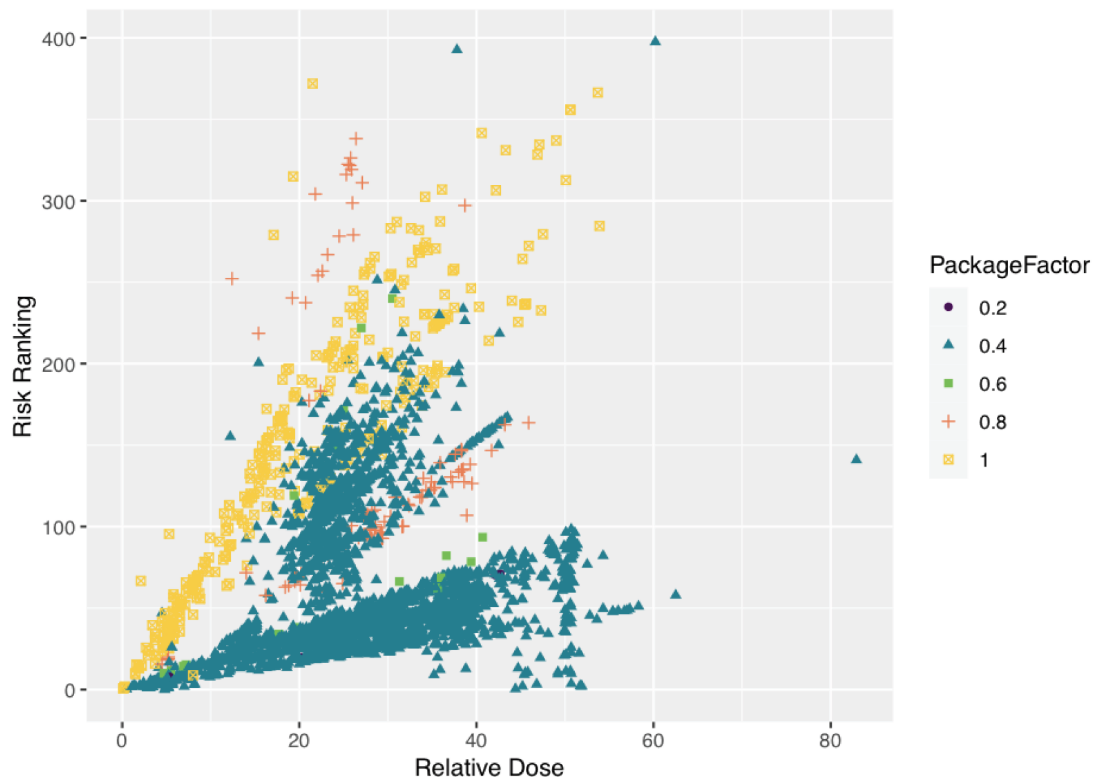

Figure 4: Scatter Plot of *RiskRanking* and *RelativeDose* grouped by *PackageFactor*.

The scatterplot of *RiskRanking* and *ReactivityIndex* by *PackageFactor* is shown in Figure 5. The *ReactivityIndex* has bands of items as it is based upon the IDC (there is a band for each IDC). Note that almost all of the *PackageFactor* = 1 (yellow box, non-standard containers) are in IDCs with the lowest *ReactivityIndex* values. Nevertheless, the *RiskRanking* values range from 0 to over 350. The *PackageFactor* grouping 0.4 (teal triangles, Hagan containers) has multiple IDCs and has a slight increasing trend of *RiskRanking versus ReactivityIndex.*

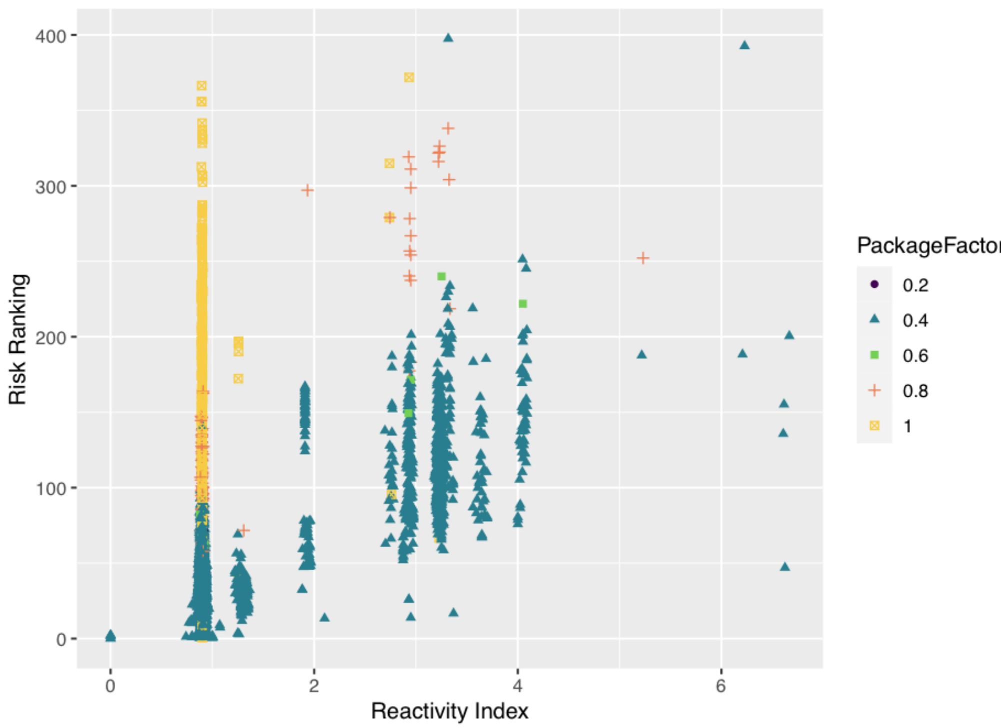

Figure 5: Scatter Plot *RiskRanking* and *ReactivityIndex* grouped by *PackageFactor*.

A 3-dimensional scatterplot for non-standard containers showing *Age* (z-axis), *RelativeDose* (x-axis) and *ReactivityIndex* (y-axis) where each point is colored by the *RiskRanking* value is shown in Figure 6. Containers with a higher *RiskRanking* are yellow, whereas containers that have a lower *RiskRanking* are blue. This scatterplot captures the interactions between the variables. Note that *ReactivityIndex* corresponds to the IDC. There is a cluster of observations (Group 1) that have a high risk ranking (over 300) a low *ReactivityIndex (< 1*) (this is difficult to see from the plot perspective), but high *RelativeDose* (34 to 54) and Age *(27-38 years)*. The containers in Group 1 are generally large with mixed material type alloyed metal or metal items. There is another group of high *RiskRanking values* (Group 2) that have lower *Age* values (19-24 years) and lower *RelativeDose (19-27)* but higher *ReactivityIndex* values (> 4.5). These container materials are generally composed of sweepings, MSE salts or larger ER salts. This figure confirms what was seen in the scatter plots, that *RelativeDose* is the main driver for *RiskRanking* for the IDCs with low *ReactivityIndex.* A high *ReactivityIndex* along with moderate *RelativeDose* and moderately high *Age* also results in a high *RiskRanking*.

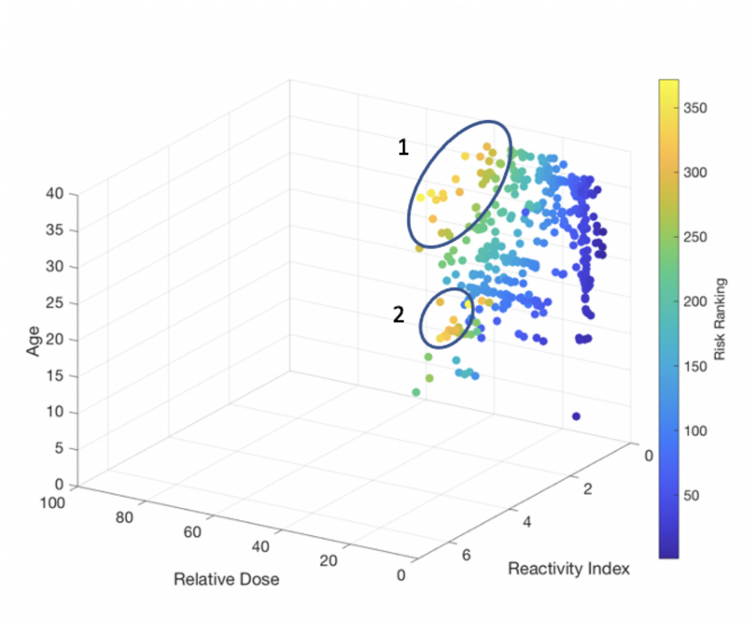

Figure 6: *Age*, *RelativeDose* and *ReactivityIndex* scatter plots with the points colored by the *RiskRanking* for non-standard containers.

Figure 7 shows a similar 3-dimensional scatterplot for the standard containers. In general, for the standard containers there is an increase in the *RiskRanking* with an increase in *RelativeDose* and *ReactivityIndex*. There are two points that stand out as having a high *RiskRanking*, both above 350, as indicated by point 1 and 2. Point 1, has a *ReactivityIndex* of 3.3, *RelativeDose* of 60 and is 19.9 years old. Point 2, has a *ReactivityIndex* of 6.2, *RelativeDose* of 37.8 and is 16.7 years old. The point 1 container is a 3QT Hagan loaded with 18g of MT83, Pu 238 R780 (sweepings). The container for this Pu 238 residue probably has a PVC bagout bag, and it is likely corroding on the inside due to the relatively high wattage (at ~0.5 watts/g it is ~9 watts), the relatively small container and the age. The point 2 container is a 5QT Hagan with 645 grams MT52 (MOX8 Pu DU Metal turnings). This item needs consideration for opening in an inert atmosphere box due to the potential pyrophoric nature of Pu metal turnings unless the comments or some other evidence indicates the material is in an hermetically sealed inner container. If it is not protected from air oxidation, the material may have oxidized by now, although that is not certain. A radiograph of this container prior to opening could indicate if the turnings are fully or partially oxidized, and could reveal if the inner container is damaged due to oxidative expansion of the contents. The integrity of these two containers could be compromised, and the risk ranking method is bringing them the attention they deserve. An extent of condition review of similar materials, container types, ages, etc. should also be considered.

Comparison of the non-standard and standard plots in Figures 6 and 7 show that all but two of the high risk containers are non-standards and that almost all of the non-standards have *ReactivityIndex* values less than or equal to 1. The standards have multiple IDCs and there is an indication that *ReactivityIndex* along with *RelativeDose* are drivers for *RiskRanking*. The graphs in this section demonstrate the value of using data visualization tools to better understand the benefits and limitations of the risk ranking method and the relative impacts of the input parameters. Other graphical methods such as star or radar plots should also be considered in future analyses, particularly for comparing subgroups.

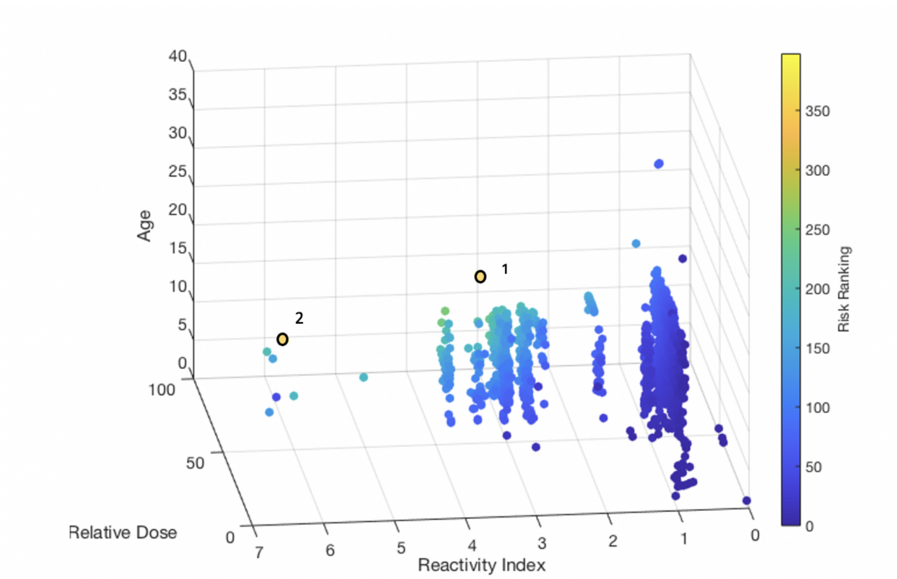

Figure 7: Age, *RelativeDose* and *ReactivityIndex* scatter plots with the points colored by the *RiskRanking* for standard containers.

2.3 How is the Risk Ranking Currently Used?

There are two primary uses of risk ranking. One is that it feeds into a container disposition prioritization scheme and the other is that it is used to track how well the Lab is doing in reducing risk. The first application is utilized in an annual planning process that involves the application of a threshold risk ranking value, such that if a container either exceeds or is expected to exceed a specified level in a given year, the container is given high priority for dispositioning that year. This process is illustrated in Figure 8. The threshold currently used is the cutoff for the “High Risk ” category. In this example, there are three containers that are in the High Risk category and are therefore identified for high-priority disposition.

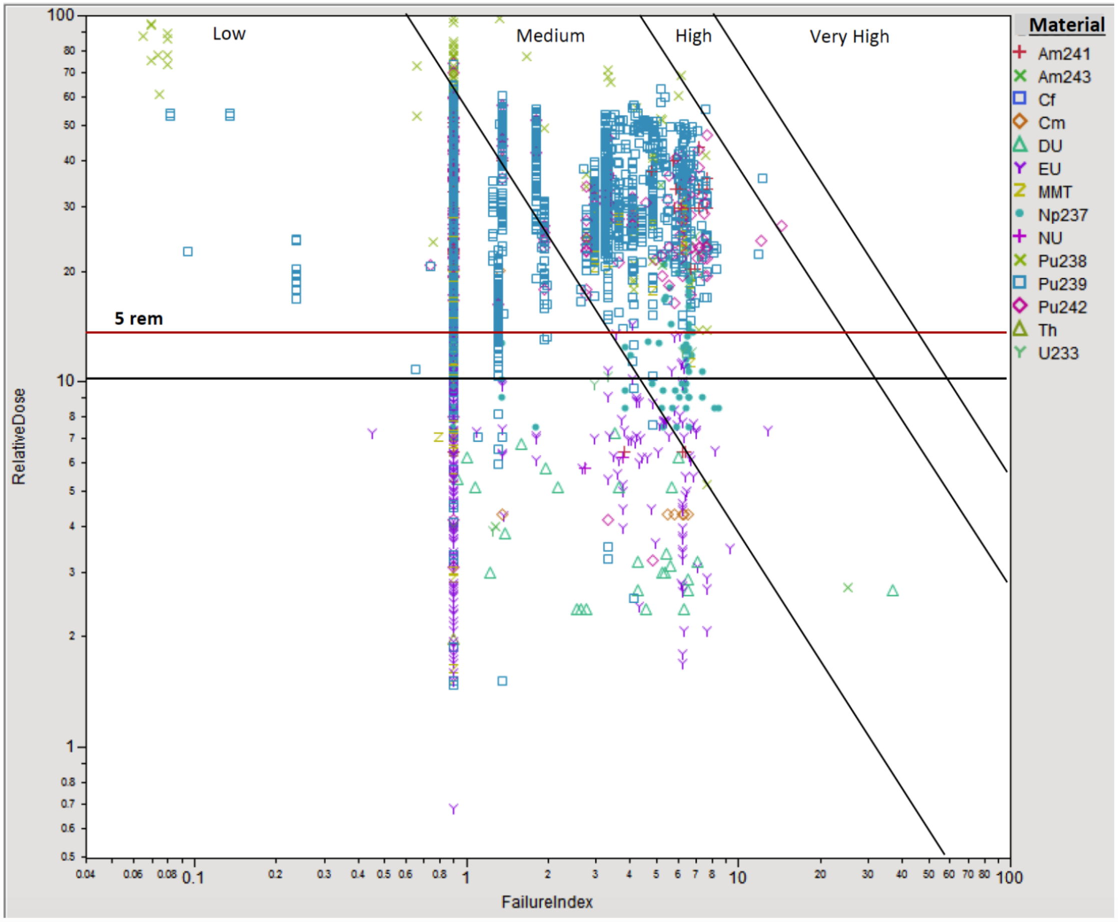

Figure 8. *FailureIndex* and *RelativeDose* are taken in combination for categorizing containers into Low, Medium, High and Very High risk. This categorization is based on a historical relative risk ranking designation and does not represent actual risk.

A common use of the risk ranking calculation is to track how well the Lab is doing in reducing risk to workers handling the containers, and to alert programmatic owners that they are responsible for repackaging containers that rise above the "Medium Risk" threshold (Abeyta, 2019). To aggregate a value to measure overall risk reduction, the sum of the *RiskRanking* values is taken over a range of containers. This is used as a measure of relative risk-reduction and reflects the reduction in the inventory of older containers with highly reactive materials in less robust containers that could contribute to a large dose if spilled.

Figure 9 illustrates an example of how relative risk was used to show repackaging progress from 2000 through 2015 and to project future relative risk (2016 to 2019) based on a particular application of repackaging resources. In this particular example, increasing package age was projected to cause relative risk for standard containers to increase if repackaging continued at a particular pace. Plots such as these can be used to evaluate repackaging plans.

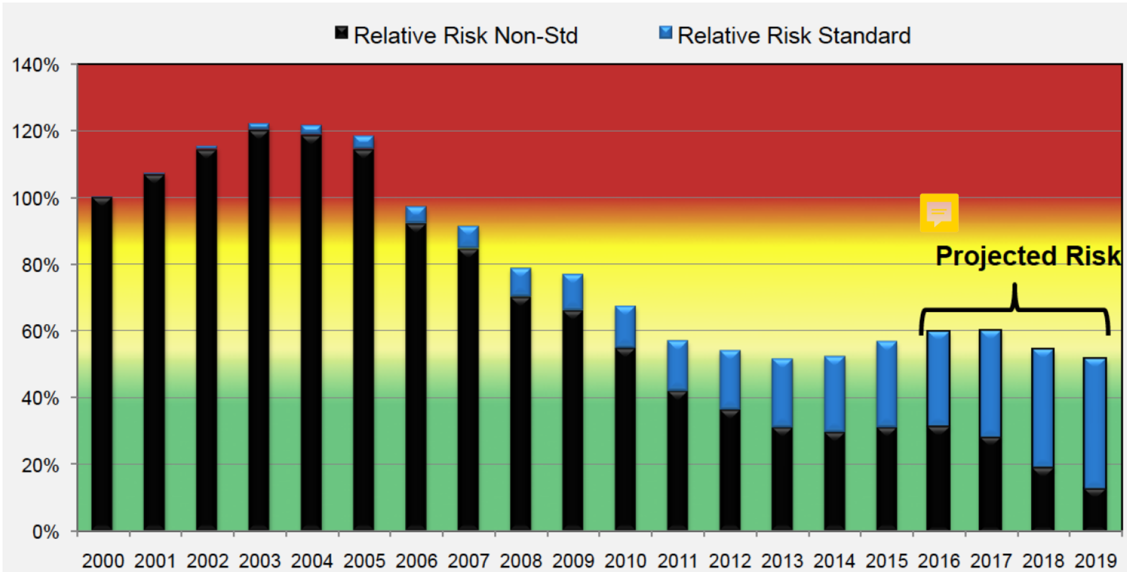

Figure 9. Bar chart illustrating relative risk ranking for standard and non-standard containers over time at the LANL Plutonium Facility. The relative risk in 2000 was set at 100% as a baseline. If repackaging resources do not keep up with the age factor, the relative risk for containers increases over time.

A variation on the risk ranking calculation is also used in conjunction with expert judgement in the LANL SAVY-4000 Field Surveillance Plan (Kaufeld, 2019). Figure 10 shows plots of dose versus item description codes (IDC). To quote from the 2020 plan,

> The parameters used to determine worst-case materials were (1) for the O-ring, those materials with the potential for a high gamma dose to the O-ring, (2) for the container body, those materials containing potentially corrosive salts and with high radiation fields of all types, and (3) for the filter, those materials with the potential to generate corrosive gases and the potential for a high gamma dose to the filter. The 12 IDC groups considered to encompass the worst-case materials are identified with blue stars. These groups were selected because they had a reasonable number of containers with the highest calculated doses and they encompassed the salt bearing residues. (The doses are estimates used for ranking purposes only and do not represent actual dose to the components.)

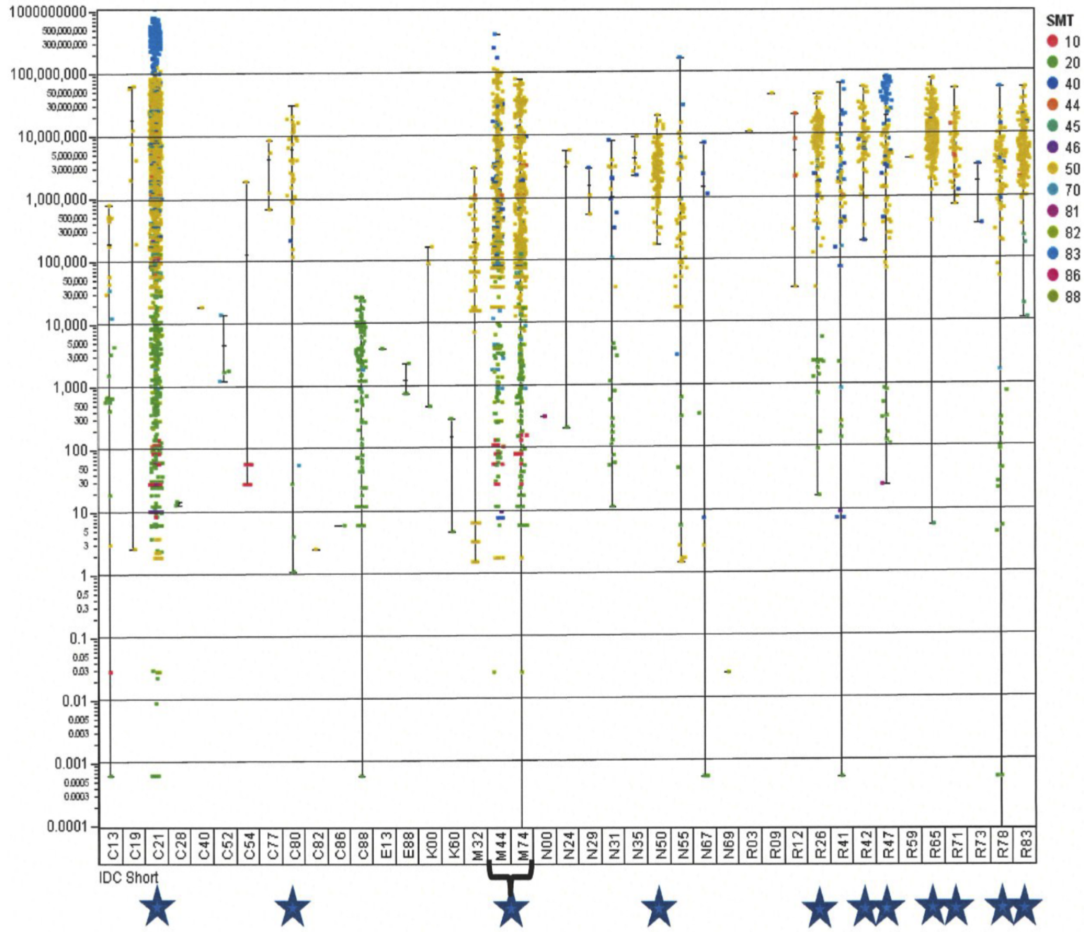

Figure 10. Calculated dose for each Item Description Code (IDC).

2.4 Prioritization of Containers Considering Worker and Programmatic Risk

The risk ranking methodology is used in conjunction with programmatic risk factors to prioritize the selection of containers for processing, discard, and shipment. Combining worker risk with factors that affect programs through vault health (e.g., capacity maintenance, capability resilience, etc.) creates a balanced strategy for long-term mission success.

The ability to prioritize items in the inventory is critical to ensuring that reducing worker risk is continually pursued and that the vault storage capability can meet the requirements for current and future missions. The prioritization calculation detailed in this section can be applied to long-term strategic goals and/or short term disposition campaigns. This flexibility allows for effective reaction or focus of item disposition planning during changing facility conditions such as paused processing lines, unavailability of discard capability, opportunities for increased throughput in specific operating lines, etc.

2.4.1 PRIORITIZATION OF ITEMS: PARAMETERS AND CALCULATION

The prioritization calculation for nuclear material items in the Los Alamos inventory uses information from the NMCA database or from the result of the item risk ranking methodology (Prochnow, 2015). All containers are assigned a priority score (PS) based on the product of the following worker (W) and programmatic risk (P) parameters using the equation: PS= A x B x C x D x E.

A. Risk ranking (W)
B. Location type (P)
C. Non-standard vs standard package (W)
D. Container size (P)
E. Packaging efficiency (P)

Each item in the inventory is assigned a value for each of the parameters (A-E) that are then entered into the PS equation. The prioritization methodology is constructed to provide relative ranking for items and not an absolute priority and/or risk value. Thus, values assigned to some parameter values were normalized to ensure model results were scaled appropriately. The normalization also ensured that a value was present for all parameters (i.e., did not allow zero values for any parameter). For example, the risk ranking parameter value was normalized within the prioritization calculation. The risk ranking values (calculated outside the prioritization calculation) are orders of magnitude beyond other parameter values to be used in the PS equation. A log scale was used to normalize the magnitude of the risk ranking input value as well as an additive factor to ensure no scores were zero.

The following defines the values that can be applied for each parameter in the calculation:

A. Risk ranking: Log([Risk Ranking value]+2); values 0.5 through 6
B. Location type attractiveness; values 2 through 10
   Based on internal priority and strategy for specific location types in the vault when establishing a disposition campaign. For example, high-mass limit, container-flexible, large locations (i.e., premium locations) score a 10 and low-criticality locations that have limited container space score a 2.
C. Non-standards vs standard; values 15 or 5
   Disposition of non-standards are a focus for the facility and receive the higher priority (value of 15)
D. Container size; actual container size in gallons
   Storage capability of larger container sizes is considered of higher priority/value
E. Packaging efficiency: 1/log(nm grams+2)
   This parameter prioritizes less efficiently packaged items for disposition. One over-the-log is used to raise items with low nuclear mass loading to higher priority. This calculation is effective at targeting items such as residues.

The output of the calculation generates a unique priority score for each item in the inventory that can then be sorted for relative prioritization. The tool output can also be combined with additional databases to enable binning of similar material types, filtering of physical forms, identify disposition path, etc. while maintaining the context of relative ranking of items.

2.4.2 *Application of Prioritization*

Whenever possible, processing, discard, and shipment of items is performed with adherence to the prioritization ranking. It is important to note that the prioritization of items is a relative ranking that is used to identify the next item to be worked when the applicable processing capability is available and that the item to be selected supports a mission need. The nuclear materials management team works to align availability of capability and capacity with feed-lists that consider the priority ranking. At times, there are focused campaigns through programs such as Material, Recycle, and Recovery (MR&R) Program that assist in dispositioning items that are unable to be processed by routine processing lines but that rank relatively high in priority. For example, the risk ranking methodology is recalculated on some frequency to identify items that have transitioned into the "high-risk" relative binning category (brought about by aging, reactivity, dose, etc.). These items are always managed as a priority over any prioritization calculation or processing feed list filtering.

The prioritization tool is used frequently to prioritize work and to analyze effects from alternative campaign scenarios. The tool is part of the ongoing continuous improvement effort to more effectively reduce risk and increase productivity at LANL. Opportunities to improve the calculation methodology, the parameters within the methodology, and its application will remain a focus within the Los Alamos long-term strategy for managing nuclear materials.

2.4.3 *Impact of RiskRanking on PS*

RiskRanking is not a driver for the calculation of PS. This is because dividing by packaging efficiency (1/log(nm grams+2)) in the PS equation is equivalent to eliminating element weight from the dose calculation in *RiskRanking* and because the PS calculation heavily weights non-standard, large containers in desirable areas (e.g., opening up storage space for weapons components). Figure 11 shows a plot of Container Priority (PS) versus *RiskRanking* for standard and non-standard containers. Standard containers are those containers with a *PackageFactor* less than 0.4 The plot shows that non-standard containers have higher *RiskRanking* than standard containers. It also shows that there is no correlation between PS and *RiskRanking*.

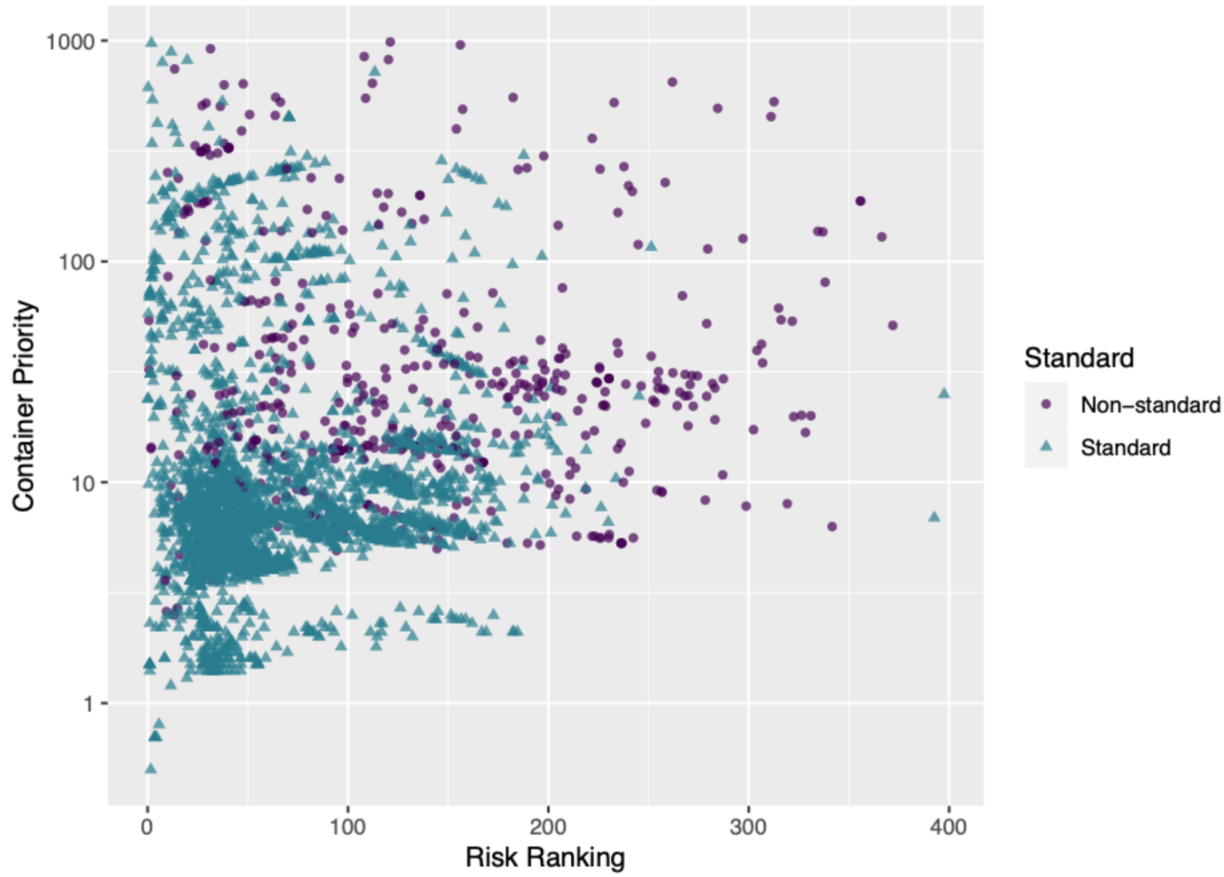

Figure 11. Container Priority (PS) versus Risk Ranking grouped by Non-standard (purple) and Standard (green) containers. Non-standards generally have higher container priority than Standards. There is no correlation between Contain Priority (PS) and *RiskRanking* for either group.

3 Case Studies illustrating positive and negative attributes of current method

Three case studies follow that are illustrative of some of the positive and negative attributes of the current risk ranking method and provide motivation for the recommended changes.

3.1 Case 1, Am-241 in Molten Salt Extraction Residues

This is an example that illustrates an issue with the current method that causes the risk for a certain class of materials to be significantly underestimated. The issue arises from the way Am-241 content is accounted for in the nuclear material accountability (NMCA) system. Up until the NMCA group reinforced the requirement to account for Am a few years back, it was recognized that some material forms (e.g., IDC=R83, MSE

Salts) had significant quantities of Am-241 that were not accounted for in the NMCA system. The language in the 474.2 Order was clarified and reinforced to say that if Am is knowingly separated as part of an operation, it must be accounted for moving forward. There have been significant efforts to account for Am since then for newly generated materials, but there are still some legacy material streams that have no Am characterization, especially if they have not been handled recently. Thus, a method more relative in nature that does not rely strictly on the "measured value" should be considered.

This issue became apparent during a recent exercise to choose some of the worst-case Hagan containers for surveillance. The timing of this exercise happened to coincide with the timing of this evaluation, and a member of the surveillance team that chose the containers for surveillance is also a member of this risk evaluation team. He noticed that a particular 5 Qt Hagan container (CAXBL129A), ~17 years old, had a particularly high heat load (estimated at 38 watts, suggesting nearly 300 g of Am-241), but the risk ranking value/category (188/Medium) appeared to be lower than the high heat load might suggest. The material in this container is a molten salt extraction (MSE salt) residue that resulted from the extraction process used to remove americium from plutonium, and these residues often have relatively large amounts and high concentrations of americium. At the time the residue was generated, Am-241 was not required by NMCA rules to be accounted for in the NMCA system. Because the risk ranking method uses data from the NMCA system, the high americium was not accounted for in the risk ranking calculation. The "missing" americium in the calculation has two major effects which, when combined, could cause the risk ranking to be seriously underestimated.

The first effect is the underestimation of the dose to the worker in the drop scenario. Americium-241 has a high dose conversion factor (DCF) of $1.52\text{X}10^{7}$ Sv CEDE/g. For perspective, the DCF for Pu-238 is $5.99\text{X}10^{7}$, and the DCF for weapons grade Pu is $3.58\text{X}10^{5}$, so the dose to the worker after a release from a dropped container with high Am-241 concentration could be very significant. If this amount of Am-241 was accounted for in the risk ranking, the value would most likely be in the high category, or possibly very high.

This case study is an example of how information from the container surveillance program should be used to update the risk ranking method. One option would be to include a section in the annual surveillance report for consideration of the appropriateness of the reactivity index values. This could work in both directions: indications that one IDC may be a problem could raise the index, but indications that an IDC thought to be problematic but shown otherwise could lower the index. This could be implemented in the FY21 surveillance plans that are currently in progress.

3.2 Case 2, Pu Metal Oxidation

Photos from the repackaging of a container named MOO1055CF that graduated to High Risk in 2019 due to age are included below in Figure 12. This container originally contained plutonium metal, primarily Pu-242, and it was packaged in an inner steel slip lid container inside a PVC plastic bag inside another steel slip lid container sometime in 1979 (a conservative estimate of the container age was 40 years based on the date the material was created in the NMCA system). The left photograph below shows the condition of the bagout bag and the inner container. The corrosion of the inner container was likely due to the degradation of the PVC plastic due to radiation and heat (plutonium metal has no chloride content, so the only available source of corrosive HCl was the PVC bag.) The right photograph shows the contents of the inner container after transfer to a new stainless steel slip lid container. The bulk of the material is plutonium oxide powder, which formed over time due to air oxidation of a plutonium metal rod or puck, leaving a relatively small amount of plutonium metal in the form of a thin metal rod. This case study demonstrates the positive attributes of the risk ranking because the method identified an item that could have failed either due to container corrosion or to oxidative expansion (Eller, 1999). It ranked highly due to its age, due to the high potential dose (highly respirable oxide), and due to the fact that it was packaged in a non-standard container. It is particularly noteworthy that the 2014 update of the risk ranking method changed the way Pu metal items were handled from a dose perspective.

The update made the more conservative assumption that Pu metal, because the containers were not hermetically sealed, transformed into the much more respirable form of Pu oxide during storage. This increased the RRF value in the dose calculation by a factor of 100, and it increased the risk ranking from 158 in 2011 (before the update) to 362 in 2013 (after the update). Because it was predicted to cross the Medium/High threshold of 377 in 2020, the material was repackaged in 2019, much earlier than it would have been otherwise. One other noteworthy point of discussion for this material is that a reactivity factor of 2.6 is used for Pu metal, and the overall range of reactivity factors is 2.6 to 7.5. Given the oxidative expansion potential for Pu metal, and the fact that it can be pyrophoric if handled in air, it is puzzling that it would have been assigned such a low value. It would seem that a reactivity factor closer to the value used for Pu metal turnings of 6.8 would be more appropriate.

Figure 12. Photographs of container MOO1055CF, a High Risk container that was repackaged in 2019. After the 2014 update of the risk ranking method, it was ranked highly due to age (~40 yr), high potential dose and the non-standard packaging configuration.

3.3 Case 3, Environmental Factors

As previously noted, one limitation of the current risk ranking method is that it does not take into account environmental factors, e.g., the effect of adjacent containers, relative humidity, temperature, etc. on container aging. This is illustrated in Figure 13. Two Hagan containers selected for surveillance in 2019 were both packaged in 2002 with nearly identical contents in the same size Hagan containers. Although the history of their storage locations is not known, the items were retrieved from different shelf locations in the same vault room at the time of surveillance. Despite being packaged with the same contents, configurations and packaging materials (presumably PVC bags), the two containers showed markedly different corrosion behavior. One of the containers had significant corrosion throughout the container giving the appearance of liquid droplets that had formed and later dried. The other container only had corrosion in the weld region, and it had bag residue on the container bottom.

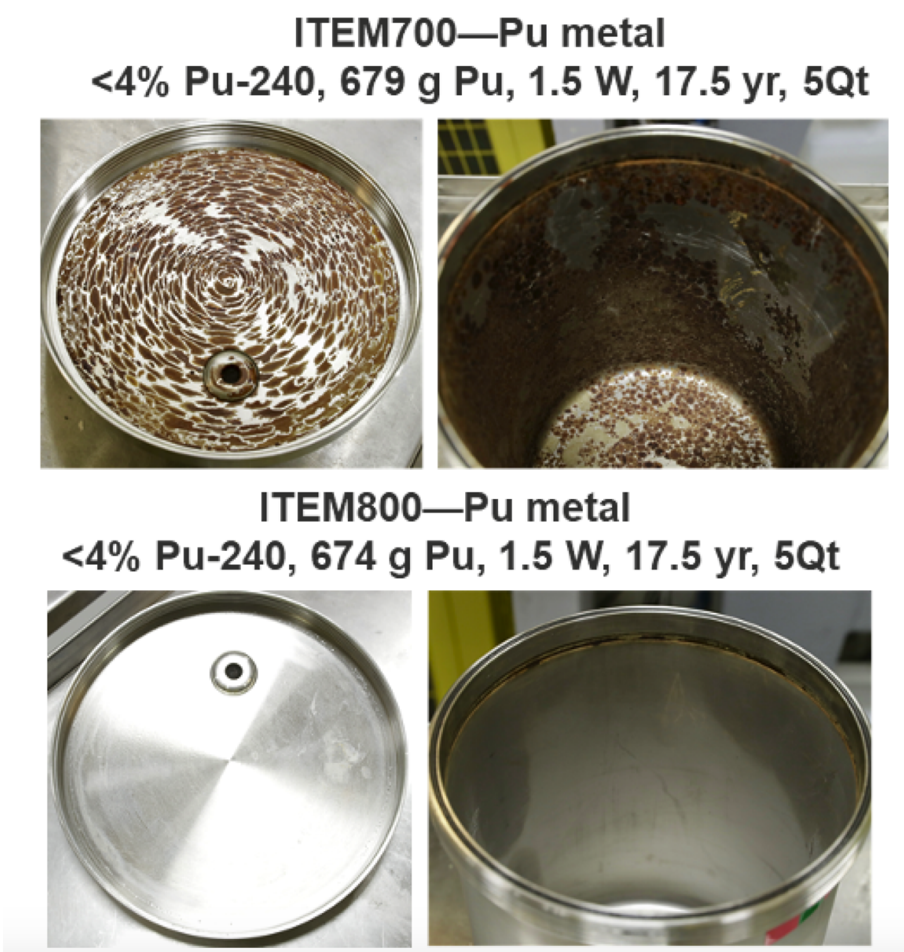

Figure 13. Comparison of corrosion behavior for ITEM700 (top) and ITEM800 (bottom) packaged with identical content for >17 years but stored in different vault locations.

4 Recommendations for Improving the Risk Ranking Methodology

Maintaining a risk-based prioritization tool for repackaging is a requirement. The current method provides a reasonable score for individual containers and their collection that allows measurement of risk reduction over time. This said, a tremendous effort at LANL relative to storage, accelerated aging studies, and expert consultation point towards a number of improvements that should be given serious consideration. The evaluation team recommends consideration of the following changes and additions to the current risk-ranking methodology.

4.1 Changes to the *FailureIndex* Calculation

4.1.1 Include PVC Bagout-bag, Heat Load, and Size in ReactivityIndex

The *FailureIndex* currently includes the *ReactivityIndex*, *RelativeDose* and *PackageFactor*. The *ReactivityIndex* is an expert judgment variable that rates from zero to three each of four hazard characteristics: pyrophoricity, corrosivity, pressure and oxidation expansion for each IDC. The sums of the ratings for each of the hazard characteristics are used to calculate the *ReactivityIndex* for each container. It is not clear that this variable adequately captures the impact of corrosion. For example, it does not specifically include the presence/absence of a PVC bagout-bag, the heat load inside the container, or the size of the container.

Recent container surveillance activities have revealed some Hagan containers where corrosion in the threads (past the o-ring seal) caused the lid to seize (Karns, 2018). In addition, corrosion is observed in the thread region of some Hagan lids. To better predict which containers are likely to exhibit this corrosion behavior, a relationship between the PVC bag degradation and the heat load, the age and the radius of the container has been suggested based on analysis of data from observations of container condition as a function of time (Smith, 2017). For those containers with a PVC bag, a bag degradation factor (BDF) has been developed. The BDF is equal to age in years times wattage/(container radius)^2. The BDF is currently being used to choose surveillance containers that are likely to be corroded (Kaufeld, 2019). The following is a quote regarding the relationship of BDF to corrosion from the most recent surveillance report (Karns, 2020):

> *The corrosion analysis for the SAVY 4000 and the Hagans indicate that high wattage and/or high BDF could result in an increased probability of having corrosion. There is considerable variability in the data and the limited number of observations (given the variability) is not sufficient for determining statistically significant results. However, with this limited data high BDF does better than wattage alone, indicating a possible interaction between wattage, age and container size.*

In 2019 an analysis of the number of Hagan containers with high BDF values was performed on the inventory at that time. It was found that there were 27 containers with BDF > 15, 42 containers with 10 <BDF <15, and 260 containers with 5 <BDF <10. To put these values in perspective, the two surveillance containers that had seized lids had BDF values of 17 and 25. The highest BDF value in the inventory for items that are likely to have a PVC bag in 2019 was ~ 40.

Surveillance data has also revealed that the MSE salt items often cause significant corrosion. Results from the surveillance program have demonstrated that the combination of high gamma dose and high heat load from the Am-241 causes the PVC bags in these containers to release significant quantities of HCl resulting in corrosion.

One way of factoring this information into the current risk ranking method would be to change the calculation of the reactivity factor or increase the existing reactivity factor based on the BDF value and americium content.

It should be noted that in the course of this evaluation it has become apparent that the estimate for the wattage for the class of materials known as MSE salts (high Am-241 concentration materials that result from the extraction of Am-241 in-growth from plutonium) is significantly and consistently a factor of 2 to 5 higher than the measured values and occasionally a factor of > 20 times higher. The measured data for wattage should be used wherever possible, and the equation for estimating wattage should be evaluated for possible improvements.

4.1.2 Incorporate Storage Conditions into the FailureIndex

Hagan drop tests have been performed on pristine containers (Karns, et al., 2016) that indicate that potential energy (storage height * container weight) and container size are key components. With a high container weight, high drop height, and particular drop orientations, it has been demonstrated that the Hagan lid comes completely off of the body after a drop, releasing significant quantities of the contents.

Currently, all containers are treated equally regardless of the container size, weight and storage height. A potential improvement of the current risk ranking could be to include in the *FailureIndex* the important factors from the drop test failure data: size, mass and the height at which a container is stored. The Hagan drop test results have been used to relieve certain vault rooms of a respirator requirement by relocating larger Hagan containers to lower shelves, thus resulting in a lower potential energy and lower probability of failure if the container is dropped.

Another limitation of the current risk ranking method is that it does not take into account environmental factors, e.g., the effect of adjacent containers, relative humidity, temperature, etc. on container aging. It may be difficult to evaluate these environmental factors, but they are likely important (see Section 3.3)

4.1.3 Regularly Update the *FailureIndex* based on Surveillance Results

The *FailureIndex* should be reviewed on a reasonable frequency (perhaps annually, or more often if a surveillance observation has significant implications for risk ranking). One possible process could be to have the Nuclear Material Storage Committee review the annual surveillance reports and plans and assess the need for updates to the risk ranking method. Another consideration would be to add a section to the annual surveillance report summarizing the lessons learned and making recommendations for which risk ranking factors should be added and/or modified. Two factors should be reviewed annually in the surveillance report: *ReactivityIndex* and *PackageFactor*. These factors could go in either direction (e.g.the *ReactivityIndex* could go down for materials more innocuous than originally thought, and the *PackageFactor* could go up because of a new issue discovered with Hagans).

One team member offered an idea that stems from his experience in the 3013 container surveillance program and is worthy of further consideration. To paraphrase, it was noted that when containers moved in or out of our "high risk" group, they were scrutinized to find out why. As different models are evaluated for LANL risk ranking, there could be substantial changes to the risk category. Understanding what causes those changes is important. Applying this to the entire inventory could be problematic, so a test bed of containers could be used. A Surveillance database containing information for ~ 120 items has been developed. The items in this database could be ranked using the current model and then compared to results using other possible models. The parameters could be adjusted to ensure that containers with failed bags and fused lids end up in the high risk category. One advantage is that the data is unclassified, so this can be done, studied and shared more readily than using classified data.

It should not be overlooked that one of the key benefits of the current risk ranking method is its ability to make a connection between real world, observable and measurable attributes (quantity, dose, reactivity, age, etc.) and the relative risk a worker faces when they enter the vault to handle a container. A member of this team once said, "It is amazing what you can find when you simply look at the data!" A corollary to this might be, "It is amazing what you can find when you simply look at the containers!" Although the surveillance program uses expert judgement to find the worst containers, much of what has been learned about container failure has arisen from vault operators seeing something unusual in the vault and reporting it. Examples include corroded TID wires, white residue accumulations (ammonium chloride) around the filter and nearby shelving surfaces, and corrosion of filters.

One way to be more intentional about this would be to harvest data that is already being collected during daily vault shelving, retrieving and surveillance. Every time a vault operator retrieves a container from the vault they perform an In Service Inspection (ISI) on all of the containers in the same location. The inspection criteria include damage, corrosion, contamination or other signs of degradation. If signs of degradation are identified, the ISI report is sent to the operations center. These reports could be collected and reviewed on a regular basis and this information could be used to complement other surveillance data. This information could also be a valuable source of measuring the health of the vault, e.g., # inspections identifying degradation/total # of inspections performed. The regular system health report prepared by the Cognizant Systems Engineer for containers collects this sort of information. Another source of information that could reveal potential container issues is the Non Destructive Assay (NDA) measurements that are performed routinely. The larger point here is that it makes sense to take best advantage of information that is collected daily by integrating across organizations and reviewing the results. This is also in line with the requirements in DOE M441.1-1 to use ALARA principles by essentially "piggy backing" container surveillance onto observations and measurements that are already being performed to meet other requirements.

4.1.4 MODIFY THE AGE COMPONENT OF THE FAILUREINDEX

Currently the age of a container is used in the *FailureIndex* for a large variety of container types. However, for example, a taped slip lid container has a shorter design life than a Hagan container and using container age by itself to calculate the *FailureIndex* is most likely giving less weight to the taped slip lid than it should have. A possible improvement is to divide *Age* by a design life. In an ideal world, all storage containers would have established design lifetimes associated with them, but only recently have design life determinations become the norm. Although a design life of 15 years has been established for SAVY containers, the actual design life may be much longer. One option would be to establish a design life for all container types based on expert judgement.

A potential complication involves the distinction between "age of" vs "use of" nuclear material storage containers. For example, there is a known population of containers that remain in vault storage that may only be handled for NMCA checks for long periods of time, and the limiting lifetime components of the container system are only being subjected to the material that is being stored inside them. If a container is being used in plant operations on a daily basis subject to wear and tear and often other various insults, then the *Age* could be scaled (perhaps based on rate of interactions or number of interactions). This scaling factor could be generated by the knowledge of the current plant operational process specific to tasks involving containers.

4.1.5 EVALUATE HOW RISK FACTORS ARE COMBINED IN THE FAILUREINDEX

In the current approach all risk factors are multiplied together, therefore, if any variable is zero, the failure index is zero. Another approach might use the square root of the sum of squares for each scaled or weighted variable. Other approaches could be considered to see if they provide a better failure ranking measure. This question was considered briefly during this evaluation, and it warrants further discussion, especially as SAVY and other standard containers age. Currently they drop out of risk ranking completely because the *PackageFactor* is zero for these containers.

Another idea that treats the inner and outer containers separately should also be considered. There is the risk to the worker of handling the inner package, and that risk is mitigated by the container. Perhaps if the ranking could be done in a way that gives two parameters: the overall risk ranking (material inside a container) and that of the inner package that has to be mitigated by radiation protection when the worker opens the package. The inner package factor could be high (R83, high gram), but if this material is inside a SAVY, the overall risk is low. The overall risk would increase gradually as the container ages. This might be accomplished by grouping the factors in the equation.

4.2 Review Parameters in Dose Calculation

The parameters in the dose calculation (isotopic composition, damage ratio, respirable release fraction, etc.) should be reviewed to ensure that they remain up-to-date and appropriate.

4.3 Review Risk Category Thresholds

The risk categories (Low, Medium and High) are based on *RiskRanking* values for the container population in 2014. There is no documentation for how they were determined. One way to avoid the terms medium risk, high risk, etc. could be to simply assign them values of 1 to 5, or A to E. This could minimize the potential to over interpret the thresholds and associate them with actual risk. In any case, the thresholds themselves should be scrutinized and broadened remembering the nature of the potentially high error bars associated with the risk ranking values.

This evaluation has also revealed that the primary avenues for reducing worker risk at LANL are based on repackaging and dispositioning materials in non-standard containers and using the threshold between medium and high risk to heavily weight the priority of containers that cross that threshold. This is a good thing as far as it goes, and has led to significant progress. However, the somewhat arbitrary nature of the way the thresholds between the risk categories were assigned could lead to an over reliance on that threshold, and the nature of the Medium Risk and High Risk terms that are used could exacerbate that problem. If this threshold continues to be used in this way, a broader look at the items that could cross the threshold from 1-5 years to 10-15 years would be appropriate. A healthy appreciation for the magnitude of the "error bars" in the risk ranking values must be maintained.

4.4 Improvements in Method for Quantifying Overall Risk

It is worth emphasizing again that the formal notion of risk entails accurate values for both the probability and consequence of a particular container's failure. Because of the great care taken in the storage of special nuclear material, data on actual failures is sparse. The *RiskRanking* provides a relative measure for containers that incorporates factors considered by SMEs to be correlated with risk. As previously described, it may not include all of the relevant factors, and just how well it correlates with risk is not known. Likewise, the change in the sum of *RiskRanking* values over time for a subset of containers is not a true measure of change in risk for those containers. In short the purpose of the risk category and risk ranking is to determine which packages pose the greatest relative risk to worker safety based on a measure that is reasonably thought to be correlated with risk.

There may be some simple measures of actual risk that should be monitored over time that are completely independent of the risk ranking method. For example, data could be compiled and analyzed on contamination incidents related to container failure, percentage of standard vs. non-standard containers, reduced external worker dose due to elimination of respirator use. This information could be used to evaluate the risk ranking model. The elimination of respirator use, where appropriate, is an indication of where we have gained confidence that the containers will protect the workers from an internal dose as they should, i.e., they are functioning as true engineered controls. General and routine reliance on respirators to protect workers from air-bourne radiation is not consistent with DOE radiation protection policy.

4.5 Consideration for Uncertainty in NMCA Database

The Am-241 issue discussed above raises a larger question about relying on data from the NMCA system to calculate relative risk ranking values. The NMCA system was developed for the purpose of tracking nuclear material to ensure every gram is in its proper place and has not been diverted. The Material Type (MT) codes, the Item Description Codes (IDC), the process status codes (PS), the gram level and wattage measurement results of fissile isotopes by NDA measurements, were all developed for accounting purposes. Other needs and requirements have capitalized on this data to track requirements related to material at risk, criticality safety, container surveillance, container prioritization and risk ranking that rely on the same kind of information.

However, the nature of the categories used to track material from an NMCA perspective is not necessarily the same categories one would choose to track container degradation or rank containers according to risk. As an example, R83 is the IDC code for MSE salt, and one might think that all MSE salts are the same. There are normal runs of the process, and then there are “failed runs,” the process residue from both runs are given the same IDC code. These materials will likely have very different chemical makeup, and could have very different corrosion behavior due to the higher amount of calcium in the failed run. The IDC code of R83 does not distinguish between these two materials. The amount of Am-241 could also vary significantly between these two materials, and it may not have been accounted for in legacy items. This could make a big difference in the corrosion behavior and the potential dose from a failed container.

The upshot of this is that results from any risk ranking based on NMCA data is only as good as the fidelity of the data itself. Delving into the details of any given item (e.g., by reading the NMCA database comments field) could put in question a high ranking for a container or raise the priority of a medium ranked container.

5 Conclusions

The purpose of this evaluation is to document the current risk ranking method, to articulate the benefits and potential limitations of this approach, and to propose changes that could lead to significant improvements in reducing worker risk and tracking risk reduction in the future.

The update of the risk ranking model in 2014 is a good example of the recognition of the importance of incorporating new information, and the improvements in the model made in 2014 were significant. After another 7 years, this evaluation has revealed some opportunities to improve the current model that are obvious and perhaps even urgent. These include the appropriate treatment of americium and measured wattage values and the creation of a process for regular integration of new information from surveillance. There are other changes that will need further consideration. These include adding parameters that are important but not accounted for in the current method (container diameter, wattage, potential energy), and adjusting the parameters that are based on engineering judgement gained from surveillance. In addition, the relative importance of every parameter should be evaluated.

Balancing worker risk with facility and programmatic risk is also critically important. A single container failure has major consequences for all three. An involved worker can have serious long term health consequences; a facility can be shut down for weeks to months; and programs that protect national security can be put at risk or fail completely. The container priority method has the same benefits and potential pitfalls as the risk ranking method, and it should undergo a similar level of scrutiny and continuous improvement on a regular basis. This evaluation has revealed that the container priority method used for repackaging and disposition has the effect of damping out the dose consequence aspect of the risk ranking method. This results from the nature of the competing needs of protecting workers from the drop of a container with a large amount of material, but at the same time eliminating large numbers of containers with a few grams each to make room for the next weapon component in the vault.

6 References

Boerigter, et al., "Risk-Based Prioritization for Plutonium Residue Holdings." 1997, LA-UR-97-0426.

Curtis, Charles B. "Criteria for Interim Storage of Plutonium Bearing Materials." 1995 LA-CP-95-0128.

Eller, P. G., et al., "Summary of Plutonium Oxide and Metal Storage Package Failures." 1999, LA-UR-99-2896.

Hoffman, et al., "Updates to Risk Ranking Algorithm for Repackaging Prioritization of LANL Nuclear Material Storage Containers." 2014, LA-UR-14-26030.

Karns, Tristan M, et al., "Worker Exposure Risk Analysis of Hagan Threaded Closure Container Drops at Los Alamos National Laboratory", LA-UR-16-24036.

Karns, Tristan M, et al., "Surveillance Report on SAVY 4000 and Hagan Nuclear Material Storage Containers for Fiscal Year 2018." 2018, LA-UR-18-31604.
Karns, Tristan M, et al., "Surveillance Report on SAVY 4000 and Hagan Nuclear Material Storage Containers Update for Fiscal Year 2020." 2020, LA-UR-21-20040.
Kaufeld, Kimberly A, et al., "Los Alamos National Laboratory SAVY-4000 Field Surveillance Plan Update for 2020." 2019, LA-UR-19-32605.
Prochnow, David A, et al., "LANL Storage Optimization Activities: Defining Vault Fullness to Inform Material Disposition Planning, Program Cost, and Out-Year Capabilities Analysis." 2015, LA-CP-15-20277.
Smith, Paul H, et al., "Risk Ranking of LANL Nuclear Material Storage Containers for Repackaging Prioritization." *Health Physics Society, The Radiation Safety Journal, Operational Radiation Safety*, vol. 92, no. suppl 2, 2007, p. S87.
Smith, Paul H, et al., "Using Container Survey Data to Assign Risk to Nuclear Material Packages." 2017 LA-UR-17-24589.
Stone, Timothy A, et al., "LANL Implementation Plan for DOE M 441.1-1, Nuclear Material Packaging Manual." 2014, LA-UR-14-25418.